Shiver Me Timbers!

Analyse von klangfarblichen Parametern der Stimme in unangenehmen Situationen

Veronika Weber, Marik Roos, Christoph Reuter
Universität Wien

universität wien

DAGA 2022 STUTTGART

Hintergrund

Im Bereich der Spracherkennung wird neben der Verarbeitung semantischer Einheiten auch die Erkennung von Emotionen immer wichtiger, nicht zuletzt im Feld der Interaktion zwischen Mensch und Maschine (Lalitha et al. 2015). In vergangenen Studien zeigte sich bereits eine Unterscheidbarkeit von Lüge und Wahrheit (Nasri et al. 2016) sowie korrekt identifiziertem Ausmaß an Zufriedenheit (Roos, 2016) in Stimmklängen, auch unabhängig von Prosodie und semantischem Kontext.

In dieser Studie sollen klangfarbliche Parameter der menschlichen Stimme in angenehmen und unangenehmen Situationen untersucht werden, um empfundenes Unwohlsein mittels Stimmklangfarbe zu operationalisieren.

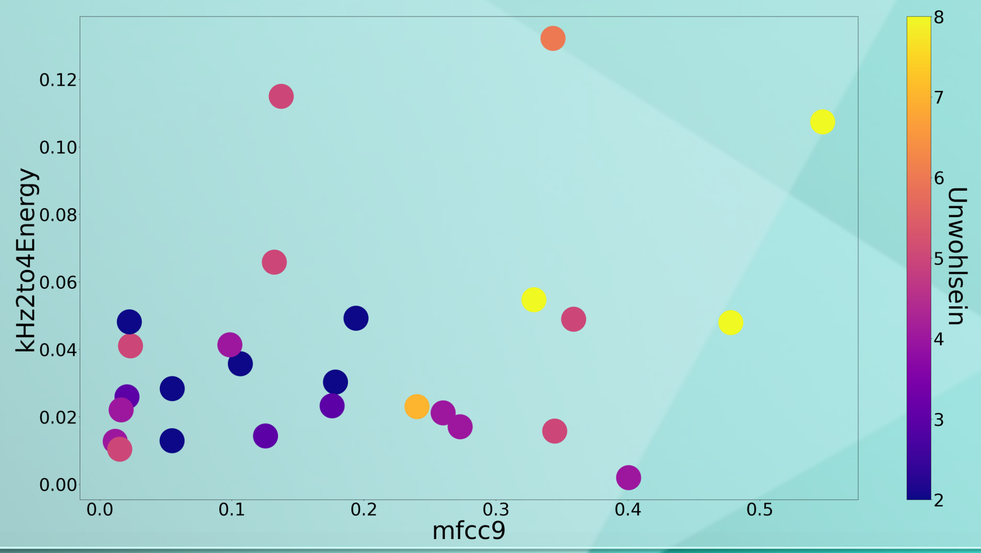
Methode

N=26 Versuchspersonen (18-38 Jahre) wurden gebeten, mehrere Texte vorzulesen. Der neutrale Text durfte von den Vpn vor der Aufnahme durchgelesen und geübt werden und diente der Ermittlung von Basislevels der Stimmklangfarbe. Die restlichen Texte sollten ohne vorherige Rezeption vorgetragen werden und beinhalteten eine Bandbreite von Statements, deren lautes Aussprechen als potentiell unangenehm estimiert wurde, z. B. Selbstlob („Ich liebe mich“), sexuell explizite Sprache, rassistische und unbekannte Wörter sowie Ich-Statements, die einen potentiellen Eingriff in die Privatsphäre der Vpn darstellen. Anschließend sollten die Vpn die Textpassagen markieren, bei deren Aussprache sie Gefühle von Unwohlsein empfunden haben.

Die Stimmaufnahmen wurden dann mittels MIRToolbox ausgewertet und einzelne Vokale sowie auch ganze Wörter und Phrasen aus den markierten Stellen auf unterschiedliche Soundfeatures analysiert.

Ergebnisse

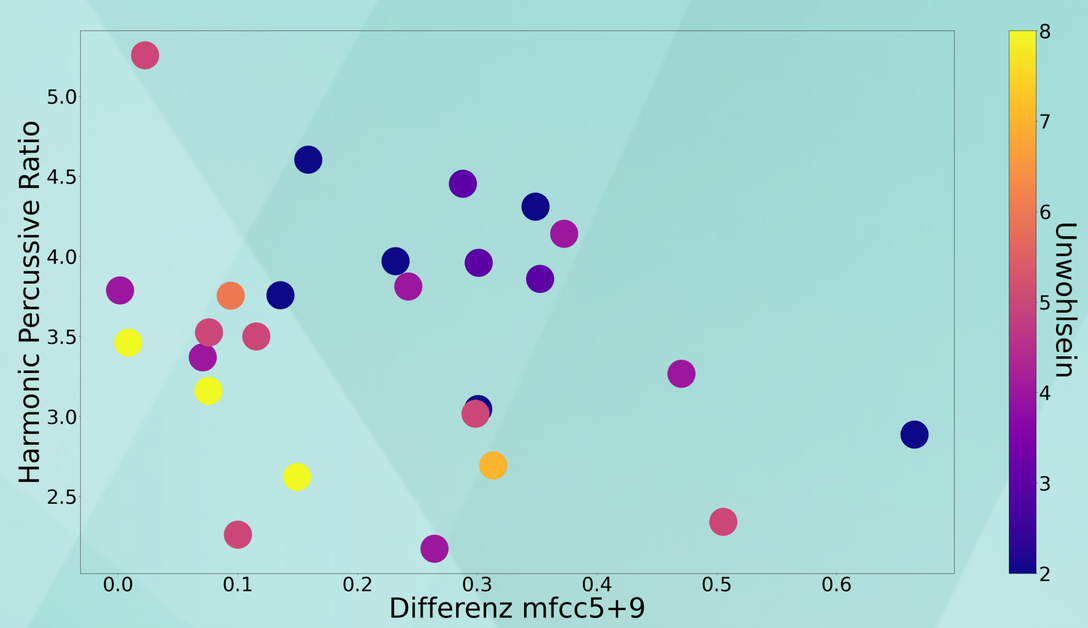
Einzelne Vokale aus dem Baseline-Text sowie aus als unangenehm markierten Phrasen zeigten in *t*-Tests keine konsistenten Effekte. Für die hellen Vokale korrelieren mit dem Ausmaß an Unwohlsein eine Erhöhung von mfcc9 (r=.642, p<.001) und 2-4kHz Energy (r=.429, p=.029). Im Regressionsmodell für „Unwohlsein“ erklären diese beiden Variablen 46% der Varianz (BF_{10}=43.229).



Gemäß der Hypothese zeigte sich eine signifikante Reduktion von mfcc3 bei unangenehmen Texten für den Vokal E (t=2.345, df=24, p=.028, d=.469) nebst Erhöhung von mfcc11 (p=.008, d=.579) und Rauhigkeit (p=.025, d=.477). Triggerwörter mit E wurden neben solchen mit I am häufigsten im Zusammenhang mit unangenehmen Gefühlen markiert. In der Analyse ganzer als Unbehagen evozierend bewerteter Textphrasen ließen sich im Vergleich zu neutralen Phrasen viele signifikante Effekte beobachten.

Parameter	t	p	Cohen's d	Parameter	t	p	Cohen's d
Envelope Flatness	6.047	< .001	1.186	Brightness High Frequency	-6.509	< .001	-1.277
Envelope Kurtosis	-6.256	< .001	-1.227	Roughness	-7.032	< .001	-1.379
mfcc1	8.619	< .001	1.690	Spectral Centroid	-3.629	0.001	-0.712
mfcc2	-3.275	0.003	-0.642	Zero Crossing Rate	-11.776	< .001	-2.309
mfcc3	-7.804	< .001	-1.530	Brightness Combination Relative	-5.730	< .001	-1.124
mfcc4	-4.242	< .001	-0.832	Band Energy Ratio High to Middle	-3.208	0.004	-0.629
mfcc5	2.409	0.024	0.473	2 to 4kHz Energy	-5.426	< .001	-1.064
mfcc6	5.427	< .001	1.064	Low Centroid Rate	6.219	< .001	1.220
mfcc7	-2.704	0.012	-0.530	Spectral Flatness	-4.640	< .001	-0.910
mfcc8	2.105	0.045	0.413	Harmonic Percussive Ratio	5.355	< .001	1.050

Als Prädiktoren für Unwohlsein eigneten sich in den Phrasen vor allem die Differenz zwischen mfcc5 und mfcc9 in Interaktion mit Harmonic-Percussive Ratio (R^2=.395, BF_{10}=26.506). Unter Einbezug von Helligkeit, Rauhigkeit, Inharmonizität, ZeroCrossingRate, LowCentroidRate sowie mfccs 2, 3 und 4 ließen sich 65.2% der Varianz für die Vorhersage unangenehmer Gefühle aufklären. Prosodische Parameter erwiesen sich hierfür als nicht relevant.



Diskussion

Die der Hypothese gemäß mit unangenehmen und angenehmen Gefühlen assoziierten mfccs 5 und 3 spielten auch für die Vorhersage des Ausmaßes an Unbehagen in dieser Studie eine Rolle. Zwar sind die Ergebnisse weniger konsistent als erhofft, jedoch könnten die Unterschiede in der Varianzaufklärung einzelner Vokale durch die Semantik der jeweiligen Wörter erklärt werden bzw. ob diese für das Unbehagen verantwortlich war oder z. B. Schwierigkeiten in der Aussprache. Die erfolgreiche Vorhersage von Unwohlsein auf der Basis klangfarblicher Parameter unter Ausschluss von Prosodie soll in zukünftigen Studien um weitere Emotionen (Freude, Trauer, Wut, Angst etc.) erweitert werden, um schließlich mittels Wahrnehmungsstudien die Verlässlichkeit der Stimmklangfarbe gegenüber Prosodie in der Identifikation von Emotionen bei Sprecher:innen zu prüfen.

Literatur

Lalitha, S., Geyasruti, D. & Narayanan, R. (2015). Emotion Detection Using MFCC and Cepstrum Features. *Procedia Computer Science, 70*, 29–35. https://doi.org/10.1016/j.procs.2015.10.020 <> Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx'07)*. Bordeaux, France, September 10-15, 2007. <> Nasri, H., Ouarda, W. & Alimi, A. M. (2016). ReLiDSS: Novel lie detection system from speech signal. *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, 1–8. https://doi.org/10.1109/aiccsa.2016.7945789 <> Roos, M., Reuter, C. (2016). Ich höre, wer Du bist! Über die Wahrnehmung geschlechtsspezifischer Persönlichkeitsmerkmale in der menschlichen Stimme. *"Akustik und Musikalische Hörwahrnehmung": 32. Jahrestagung der Deutschen Gesellschaft für Musikpsychologie (DGM)*, Wien, Österreich.